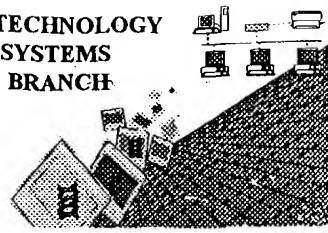


BIOTECHNOLOGY  
SYSTEMS  
BRANCH



**RAW SEQUENCE LISTING**  
**ERROR REPORT**

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/752,957D  
Source: IFW/b  
Date Processed by STIC: 3/1/04

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 703-308-4212; FAX: 703-308-4221

Effective 12/13/03: TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 4.1 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker/chkr41note.htm>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebc/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. Hand Carry directly to (EFFECTIVE 12/01/03):  
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Revised 10/08/03

## Raw Sequence Listing Error Summary

<u>ERROR DETECTED</u>	<u>SUGGESTED CORRECTION</u>	<u>SERIAL NUMBER:</u> <u>09/752,957D</u>
<b>ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE</b>		
1 <input checked="" type="checkbox"/> Wrapped Nucleic Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."	
2 <input checked="" type="checkbox"/> Invalid Line Length	The rules require that a line <b>not exceed</b> 72 characters in length. This includes white spaces.	
3 <input checked="" type="checkbox"/> Misaligned Amino Numbering	The numbering under each 5 <sup>th</sup> amino acid is misaligned. Do <b>not</b> use tab codes between numbers; use <b>space characters</b> , instead.	
4 <input checked="" type="checkbox"/> Non-ASCII	The submitted file was <b>not</b> saved in ASCII(DOS) text, as required by the Sequence Rules. <b>Please ensure your subsequent submission is saved in ASCII text.</b>	
5 <input checked="" type="checkbox"/> Variable Length	Sequence(s) _____ contain n's or Xaa's representing more than one residue. <b>Per Sequence Rules, each n or Xaa can only represent a single residue.</b> Please present the <b>maximum</b> number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.	
6 <input checked="" type="checkbox"/> PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. <b>This applies to the mandatory &lt;220&gt;-&lt;223&gt; sections for Artificial or Unknown sequences.</b>	
7 <input checked="" type="checkbox"/> Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for <b>each</b> skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped	
	Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to <b>include</b> the skipped sequences.	
8 <input checked="" type="checkbox"/> Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If <b>intentional</b> , please insert the following lines for <b>each</b> skipped sequence. <210> sequence id number <400> sequence id number 000	
9 <input checked="" type="checkbox"/> Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is <b>MANDATORY</b> if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.	
10 <input checked="" type="checkbox"/> Invalid <213> Response	Per 1.823 of Sequence Rules, the <b>only valid</b> <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is <b>required</b> when <213> response is Unknown or is Artificial Sequence	
11 <input checked="" type="checkbox"/> Use of <220>	Sequence(s) _____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is <b>MANDATORY</b> if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)	
12 <input checked="" type="checkbox"/> PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.	
13 <input checked="" type="checkbox"/> Misuse of n/Xaa	"n" can <b>only</b> represent a single <u>nucleotide</u> ; "Xaa" can <b>only</b> represent a single <u>amino acid</u>	



IFW16

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/752,957D

DATE: 03/01/2004

TIME: 07:12:55

Input Set : A:\251410-1010.ST25.txt

Output Set: N:\CRF4\03012004\I752957D.raw

3 <110> APPLICANT: Shuan, David  
 5 <120> TITLE OF INVENTION: Yeast with High Biotin-Productivity and the Preparation  
 Method

6 Thereof

8 &lt;130&gt; FILE REFERENCE: 251410-1010

10 &lt;140&gt; CURRENT APPLICATION NUMBER: 09/752,957D

11 &lt;141&gt; CURRENT FILING DATE: 2001-01-02

13 &lt;160&gt; NUMBER OF SEQ ID NOS: 5

15 &lt;170&gt; SOFTWARE: PatentIn version 3.2

17 &lt;210&gt; SEQ ID NO: 1

18 &lt;211&gt; LENGTH: 1188

19 &lt;212&gt; TYPE: DNA

20 &lt;213&gt; ORGANISM: Nucleic Acid

22 &lt;400&gt; SEQUENCE: 1

23 atgtcgaaaa tattgactgc tattagtcgt ccgattgctc tttccacttc tagagtagct

60

25 tctaggggcta ctctggcaac aggtgtact gctgctgcgg agatcttggaa agatgtgttc

120

27 acggaacaaa tggagaagaat ggcttcacag gagaagaagc caaaccatt ggaatatgc

180

29 ttgtcagtga agacaccgt caacacctgg accaaagaag aaattaaagc tatatatgac

240

31 acaccactca tggacttgat gcactatgct caggtgcaac acagaagggtt ccaacaacct

300

33 tcagagggttc aattgtgcac tcttatgaat atcaaaactg gtgggtgtac cgaggactgt

360

35 aagtactgtg cccaatcaca gcgttacaac actgggtgtca aggctgaaag aatcatccaa

420

37 gttgatgagg tgattgaagc tgcaaaggag gcaaaggcca atggatctac aagggtctgt

480

39 atgggtctg cttggagaga gatgaaagggt agaaagtcaa acttgaagaa aatcaaagag

540

41 atgatcactg ctgtccatga ccttggaatg gagagttgtc tcaccctggg aatgggtgtat

600

43 aaagaccaag ccactgaatt gaaaagtgtc ggggtgacgg cgtacaacca taacattgtat

660

45 acttacaagg aacactatcc aaaggtgatc tccacaagaa gctttatgtat tagattgaaa

720

47 acattcaaaa acgttcaagg atctggatta aaggcatgca caggtgttat tcttggctt

780

49 ggtgagaccc aaggaggaccc tggatcttc ctctcacactt tggccacaat ggatcagcat

840

51 ccagagtctc ttccaaatcaa cagactggtc ccaatcaagg gcacgccaat gtatgaagaa

900

53 gttaaagaaca agcaaggatgtg agttgtatgat attgtcagaa ccattgtctac tgcaagattg

960

55 gtcataccaa cgttctattat cagattggct gcaggaagat atacaatgaa agaggcagaa

1020

57 caggtgtatgt gtttcatggc tggatgtat gccatcttca caggtaaagaa aatgctcaca

1080

59 acaatgtgtaa acggctggaa tgaggataaaa gccatgttgg ctaaatgggg tctgaaaccca

1140

61 atggagatgtt tcaaatacaa accaaggggat gttgcattcg gtgtctga

1188

64 &lt;210&gt; SEQ ID NO: 2

65 &lt;211&gt; LENGTH: 30

66 &lt;212&gt; TYPE: DNA

67 &lt;213&gt; ORGANISM: Gene

same error

69 &lt;400&gt; SEQUENCE: 2

70 gaaagtgcac tcaagatctg tcgtacttaa

30

73 &lt;210&gt; SEQ ID NO: 3

74 &lt;211&gt; LENGTH: 21

75 &lt;212&gt; TYPE: DNA

76 &lt;213&gt; ORGANISM: Gene

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/752,957D

DATE: 03/01/2004

TIME: 07:12:55

Input Set : A:\251410-1010.ST25.txt

Output Set: N:\CRF4\03012004\I752957D.raw

78 <400> SEQUENCE: 3  
 79 ccgcagttaa atcgacaact g  
 82 <210> SEQ ID NO: 4  
 83 <211> LENGTH: 23  
 84 <212> TYPE: DNA  
 85 <213> ORGANISM: Amino Acid  
 88 <220> FEATURE:  
 89 <221> NAME/KEY: misc\_feature  
 90 <222> LOCATION: (4)..(4)  
 91 <223> OTHER INFORMATION: n is a, c, g, or t  
 93 <220> FEATURE:  
 94 <221> NAME/KEY: misc\_feature  
 95 <222> LOCATION: (6)..(6)  
 96 <223> OTHER INFORMATION: n is a, c, g, or t  
 98 <220> FEATURE:  
 99 <221> NAME/KEY: misc\_feature  
 100 <222> LOCATION: (18)..(18)  
 101 <223> OTHER INFORMATION: n is a, c, g, or t  
 103 <400> SEQUENCE: 4

21

1) invalid <213> response  
 2) This sequence is not an amino acid  
 sequence.

OK

&gt; 104 tgtnncngarg aytgyaanta ttg

23

107 <210> SEQ ID NO: 5  
 108 <211> LENGTH: 20  
 109 <212> TYPE: DNA  
 110 <213> ORGANISM: Amino Acid  
 113 <220> FEATURE:  
 114 <221> NAME/KEY: misc\_feature  
 115 <222> LOCATION: (6)..(6)  
 116 <223> OTHER INFORMATION: n is a, c, g, or t  
 118 <220> FEATURE:  
 119 <221> NAME/KEY: misc\_feature  
 120 <222> LOCATION: (8)..(8)  
 121 <223> OTHER INFORMATION: n is a, c, g, or t  
 123 <400> SEQUENCE: 5

Same error as above

OK

&gt; 124 gtrtcnanrt trtggttgta

20

**RAW SEQUENCE LISTING ERROR SUMMARY**  
PATENT APPLICATION: US/09/752,957D

DATE: 03/01/2004  
TIME: 07:12:56

Input Set : A:\251410-1010.ST25.txt  
Output Set: N:\CRF4\03012004\I752957D.raw

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; N Pos. 4,6,18

Seq#:5; N Pos. 6,8

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/752,957D

DATE: 03/01/2004

TIME: 07:12:56

Input Set : A:\251410-1010.ST25.txt

Output Set: N:\CRF4\03012004\I752957D.raw

L:104 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:0  
L:124 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:0